BIOLOGY OF REPRODUCTION **69,** 469–474 (2003)
Published online before print 2 April 2003.
DOI 10.1095/biolreprod.102.015172

Serial Analysis of Gene Expression in Turkey Sperm Storage Tubules in the Presence and Absence of Resident Sperm

Ezhou L. Long,[2,3] Tad S. Sonstegard,[4] Julie A. Long,[3] Curtis P. Van Tassell,[4] and Kurt A. Zuelke[1,3]

Biotechnology and Germplasm Laboratory[3] and Bovine Functional Genomics Laboratory,[4] Animal and Natural Resources Institute, USDA-ARS, Beltsville, Maryland 20705

ABSTRACT

Turkey sperm lose viability within 8–18 h when stored as liquid semen using current methods and extenders. In contrast, turkey hens maintain viable, fertile sperm in their sperm storage tubules (SST) for 45 or more days following a single insemination. Our long-term objectives are to identify and characterize differentially expressed genes that may underlie this prolonged sperm storage and then use this information to develop improved methods for storing liquid turkey semen. We employed serial analysis of gene expression (SAGE) to compare gene expression patterns in turkey SST recovered from hens after artificial insemination (AI) with extended semen (sperm AI) or extender alone (control AI). We constructed two separate SAGE libraries with SST RNA obtained from sperm and control AI hens. We used these libraries to generate 95 325 ten-base pair SAGE tags. These 95 325 tags represented 27 430 unique genes. The sperm and control AI libraries contained 47 663 and 47 662 tags representing 18 030 and 19 101 putative unique transcripts, respectively. Approximately 1% of these putative unique genes were differentially expressed ($P < 0.05$) between treatments. Tentative annotations were ascribed to the SAGE tag nucleotide sequences by comparing them against publicly available SAGE tag and cDNA sequence databases. Based on its SAGE tag nucleotide sequence, we cloned a partial turkey *avidin* cDNA and confirmed its up-regulation in the sperm AI SST. The bioinformatics and experimental procedures employed to clone turkey *avidin* and confirm its differential expression represent a useful paradigm for analyzing SAGE tag data from relatively uncharacterized model systems.

gene regulation, oviduct, sperm, sperm motility and transport

INTRODUCTION

Because of prolonged sperm storage in their distal oviduct, female poultry can produce fertile eggs for several weeks following a single natural mating or artificial insemination (AI) [1, 2]. Oviductal sperm storage is a feature common to all avian and a wide range of other species, including newt [3], snakes [4, 5], frogs [6], lizards [7], and turtles [8]. Fertile sperm can be maintained in the turkey hen oviduct from 45 to 112 days following either AI or natural mating [2]. Avian sperm are stored in specialized sperm storage tubules (SST) localized within the luminal mucosal epithelium of the uterovaginal junction (UVJ) [1, 9]. Turkey SST are comprised primarily of tall columnar, nonciliated, nonsecretory epithelial cells that appear as bud-like invaginations of the UVJ mucosal epithelium [10, 11]. In marked contrast to the prolonged sperm storage observed in SST, current best practices for handling and storing liquid turkey semen in vitro only maintain sperm viability for approximately 8–18 h [12, 13].

The spermatozoa are closely apposed to the SST luminal epithelium, and numerous physiological interactions are likely both between and within these separate cell types during prolonged sperm storage in turkeys [1, 9]. Sperm localized within the SST are typically immotile. Thus, general mechanisms that reversibly suppress sperm respiration and motility, stabilize sperm cell membranes and enzyme systems, and suppress sperm immunogenicity within the SST have been proposed to mediate prolonged sperm storage [14–17]. Turkey intraluminal SST epithelial cells contain relatively high levels of carbonic anhydrase, suggesting that intraluminal SST pH might play a significant role in modulating sperm motility during oviductal sperm storage and transport [18]. The presence of progesterone and estrogen receptors in SST of laying hens, but not in immature chicks, suggests that steroid hormones could also play a role in prolonged sperm storage [19]. Elucidating and characterizing the molecular mechanisms that enable prolonged sperm storage would assist in developing more efficient methods for preserving turkey semen in vitro. Because virtually all turkey breeders in the United States use AI, improving sperm viability during short-term liquid semen handling and storage would increase reproductive management options and reduce turkey production costs.

Based on the close apposition between sperm and SST epithelium combined with the physiological modifications that occur in the sperm (e.g., decreased motility) within the SST, we hypothesized that specific and characteristic gene expression events within the turkey SST enable and regulate prolonged sperm storage. An objective of the present study was to perform comparative serial analysis of gene expression (SAGE) between turkey SST with or without resident sperm. A relatively new research strategy, SAGE enables transcriptome-wide qualitative and quantitative analysis of gene expression within tissues during discreet physiological states [20, 21]. We constructed two turkey SAGE libraries from SST total RNA obtained 48 h after AI with either extended semen (sperm AI) or extender alone (no sperm; control AI) and then compared SAGE tag frequencies between these two libraries to generate a list of putative differentially expressed genes in the SST epithelial cells. We then developed an analytical and bioinformatics paradigm to obtain tentative sequence annotations for the

[1]Correspondence: Kurt A. Zuelke, Biotechnology and Germplasm Laboratory, Animal and Natural Resources Institute, USDA-ARS, Bldg 200, Rm 124A, BARC-East, Beltsville, MD 20705. FAX: 301 504 5123; e-mail: kzuelke@anri.barc.usda.gov
[2]Current address: National Cancer Institute, Center for Cancer Research, Mammary Biology and Tumorigenesis Laboratory, Bldg 10, Rm 5847, Bethesda, MD 20892-1402.

Received: 31 December 2002.
First decision: 21 January 2003.
Accepted: 26 March 2003.

ISSN: 0006-3363. http://www.biolreprod.org

genes represented by the turkey SAGE tag frequency and nucleotide sequence data. We identified a differentially sampled SAGE tag corresponding to turkey *avidin* and confirmed its differential expression via quantitative real-time reverse transcriptase-polymerase chain reaction (RT-PCR) to validate this paradigm.

MATERIALS AND METHODS

Animals

Large White BUTA (British United Turkeys of America, Lewisburg, WV) breeder turkeys were maintained under standard management conditions and photostimulated on a daily basis with a 14L:10D photoperiod. The hens were kept individually in cages, and the toms were kept in groups of 8–10 in floor pens. Feed and water were provided ad libitum. Semen was collected by abdominal massage and diluted 1:1 with Beltsville Poultry Semen Extender (BPSE), and the sperm concentration was determined [11]. For all hens, AI was performed using 1.5×10^8 sperm/hen. To generate SST samples for construction of the control library (no sperm), three randomly selected hens were sham-inseminated using only BPSE without sperm. Typically, between 65% and 80% of the SST contain sperm at 48 h after AI under these conditions [11, 22].

Sample Collection and Total RNA Extraction

Hens selected for SST collection within each treatment group had all laid an egg on the morning of collection. Hens were killed by cervical dislocation 48 h after insemination to achieve maximal filling of the SST [9]. The oviduct was immediately removed from the bird and quickly trimmed free of connective tissue surrounding the vagina and uterus to reveal the UVJ, and the SST were recovered by scraping the mucosal epithelium of the UVJ with a scalpel blade [23]. Samples were snap-frozen in liquid nitrogen and maintained at −70°C until RNA extraction. Total RNA was isolated using the Totally RNA Kit (Ambion, Austin, TX) according to the manufacturer's instructions. Briefly, approximately 80 mg of SST mucosal epithelium were lysed in 800 μl of denaturing solution using a Tissuemizer Mark II T25 homogenizer (Tekmar Company, Cincinnati, OH). After phenol/chloroform extraction, total RNA was precipitated with an equal volume of isopropanol and washed with 70% (v/v) ethanol. The RNA pellets were solubilized in 0.1% (v/v)-diethyl pyrocarbonate-treated dH_2O and stored at −70°C. Integrity of the total RNA was assessed by 1% (w/v) agarose-formaldehyde gel electrophoresis.

Construction of SAGE Libraries

Aliquots of total RNA from the SST of three hens (~3.5 μg RNA/bird) were pooled within each treatment group before constructing the SAGE libraries with the I-SAGE kit (Invitrogen, Carlsbad, CA). The protocol of this kit is based on the original SAGE methodology [20; see also http://www.sagenet.org]. The anchoring and tagging restriction enzymes used were *Nla*III and *Bsm*FI, respectively. Following PCR amplification, di-tags (26mers) were released using *Nla*III and separated via 12% (w/v) PAGE. The 26mer di-tags were purified from the polyacrylamide gel, ligated to form concatemers, and size fractionated via 8% (w/v) PAGE. Three size ranges of di-tag concatemers (300–500, 500–800, and >800 base pairs [bp]) were isolated. The 500- to 800-bp fraction was ligated into *Sph*I-linearized pZErO-1 vector (Invitrogen) to construct the present libraries. Ligation products were transformed into One Shot TOP10 Electrocomp cells (Invitrogen) by electroporation and cultured overnight at 37°C on low-salt LB agar plates containing 50 μg/ml of zeocin. The SAGE tag inserts were amplified by inoculating individual colonies into separate wells of a 96-well PCR plate containing 25 μl of a PCR reaction mix consisting of 2.5 μl of 10× PCR Buffer (200 mM Tris-HCl [pH 8.4] and 500 mM KCl), 1.25 μl of dimethyl sulfoxide, 500 μM dNTP, 1.7 μM $MgCl_2$, 0.2 μM M13 forward (5′-CCCAGTCACGACGTTGTAAAACG-3′) and reverse (5′-AGCGGATAACAATTTCACACAGG-3′) primers, and 1 U of Platinum Taq DNA polymerase (Invitrogen). The thermocycling profile was denaturation at 94°C for 30 sec, annealing at 55°C for 30 sec, and extension at 70°C for 90 sec for 27 cycles. The PCR products were purified with a 96-well Montage PCR cleanup kit (Millipore, Bedford, MA) and recovered in 50 μl of nuclease-free water. For sequence analysis, 2 μl of the purified PCR product template were transferred to a 384-well plate containing 0.5 μl of Big-Dye v. 2.0 (Applied Biosystems, Foster City, CA), 1.5 μl of Big-Dye extender (Sigma, St. Louis, MO), and 3.2 pmol of SP6 primer (5′-ATTTAGGTGACACTATAG-3′) and were then amplified using standard thermocycling conditions as recommended by the manufacturer of Big-Dye. Reaction products were precipitated with four volumes of 70% (v/v) isopropanol and washed with four volumes of 70% (v/v) ethanol. After drying, reaction products were resuspended in 25 μl of Hi-Di Formamide (Applied Biosystems), denatured at 95°C for 5 min, and analyzed on an ABI-3700 automated DNA analyzer (Applied Biosystems).

Processing and Analysis of SAGE Tag Sequences

Sequence quality assessment and trimming were performed with phred v0.980904.e. Vector sequence was identified and trimmed using ''cross_match'' with the ''-minscore 18'' and ''-minmatch 12'' options. Sequence information in the processed trace files was converted into text files, and the SAGE tags were extracted and quantified using SAGE 2000 software version 4.12 (http://www.sagenet.org/Software/software2000.htm). Tag nucleotide sequences and frequency data were then outputted to MS Access (Microsoft, Redmond, WA) database files for subsequent analyses. For determining differential expression, tag frequencies between both SAGE libraries were analyzed for significance ($P < 0.05$) using tools in the SAGE 2000 software based on a chi-square analysis combined with Monte-Carlo simulations [24].

Annotation of SAGE Tags

Because the *Nla*III-recognition sequence CATG lies immediately 5′ of the SAGE tag sequence, CATG was appended 5′ to each SAGE tag to yield 14-bp tags for use in database comparisons [20]. Three primary-sequence databases were used to ascribe tentative annotation to the turkey-derived SAGE tags. First, tag sequences were compared with an existing database of human SAGE tags (human SAGEmap database; National Center for Biotechnology Information [NCBI], Bethesda, MD; ftp://ftp.ncbi.nih.gov/pub/sage/map) [25]. Turkey-derived SAGE tags were then also compared against cDNA sequence information in the chicken gene index (GgGI, version 4) available from The Institute of Genome Research (TIGR; Rockville, MD; http://www.tigr.org/tdb/tgi/gggi) [26] and a chicken expressed sequence tag (EST) database (http://www.chick.umist.ac.uk/cgi-bin/chicken_database.cgi; version 11/06/02) available from the Biotechnology and Biological Sciences Research Council (BBSRC; Roslin, U.K.). Database comparisons were performed using BLAST algorithms [27] (http://www.ncbi.nlm.nih.gov/BLAST/) and required a perfect 14-base match for inclusion in the final data set. Tentative annotation of the turkey SAGE tag representing *avidin* was also confirmed by BLAST analysis of this tag with the GenBank nonredundant (nr) database (NCBI; http://www.ncbi.nlm.gov/). To increase the likelihood that these tentative SAGE tag annotations could be ascribed to a defined functional category (e.g., metabolism, cytoskeletal protein, etc.), initial database matches were filtered using the following criteria: 1) those that aligned with differentially expressed ($P < 0.05$) SST SAGE tags, 2) those that were represented in a library five or more times, 3) those that were linked to a Unigene or TIGR Tentative Consensus (TC) identifier (i.e., tentative contiguous EST assemblies representing a predicted gene annotation), and 4) those that were annotated with a gene designation rather than simply an EST clone identifier (i.e., non-EST annotations). These search criteria yielded the most consistent gene annotations during preliminary analyses of turkey SST and swine embryo SAGE library data sets in our laboratory (data not shown).

5′-Rapid Amplification of cDNA Ends

Capture of cDNA sequence information corresponding to a SAGE tag tentatively identified as turkey *avidin* was achieved using the 5′-rapid amplification of cDNA ends (RACE) System (Invitrogen). First-strand cDNA synthesis was primed with an aliquot of the sperm-treated SST RNA used previously for SAGE library construction and the primer (5′-IIIIIGCAGCAGCCACATG-3′) consisting of 10 nucleotides complementary to the tag sequence flanked by CATG (*Nla*III-recognition site) and five inosine nucleotides to aid annealing of primer through nonspecific base-pairing [28]. First-strand cDNA synthesis products were dC-tailed and amplified by PCR with an antisense-primer designed from chicken *avidin* sequence and complementary to the 14-bp tag sequence (5′-GCAGCAGCCACATGGTCTTC-3′) and the abridged anchor primer provided in the RACE kit. Products were purified, sequenced, and identified by BLAST analysis as described above.

Real-Time PCR

Real-time PCR was performed using the QuantiTect SYBR Green RT-PCR kit (Qiagen, Valencia, CA) following the manufacturer's instruction. First-strand cDNA was synthesized from total RNA with the oligo-dT(12–18) primer (Invitrogen) and PCR-amplified in a DNA Engine Opticon Continuous Fluorescence Detector (MJ Research, Waltham, MA) for as many as 60 cycles using primers specific to turkey *avidin* (sense, 5′-GGCTCCAACATGACCATC-3′; antisense, 5′-GGTGGACTCTGAAAACTTCC-3′). Primers (sense, 5′-CCATGTTTGTGATGGGTGTC-3′; antisense, 5′CTCCACAATGCCAAAGTTGT-3′) specific for turkey *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) were used as an internal control for amplification. The *GAPDH* primers were designed based on the partial turkey GAPDH sequence (GenBank Accession no. U94327). Specificity of PCR products was confirmed by melting-curve analysis and gel electrophoresis. Real-time PCR was repeated in triplicate for each sample within an individual experiment. The threshold cycle (C_T) was the cycle where increasing fluorescent product was first detectable. Linear standard curves (C_T vs. logarithm of cDNA concentration) were plotted to calculate the amplification efficiency (AE) of each primer pair by using a logarithmic dilution of the cDNA mix reverse transcribed from the control AI RNA [29]. The AE was calculated using $10^{(-1/\text{slope of standard curve})}$. The fold difference in *avidin* expression between two libraries was calculated as $(AE^{\Delta C_T(\text{Control – Sperm AI})\text{Avidin}})/AE^{\Delta C_T(\text{Control – Sperm AI})\text{GAPDH}}$, in which $^{\Delta C_T(\text{Control – Sperm AI})x}$ indicates the difference in C_T of a particular gene X between two libraries [29].

RESULTS

Construction and Comparison Between AI and Control Insemination SAGE Libraries

Two turkey SST SAGE libraries were constructed and used to generate a total of 95 325 tags. The control and sperm AI libraries contained 47 662 and 47 663 tags representing 19 101 and 18 030 putative unique transcripts, respectively. Complete tag sequences and frequencies from these libraries have been deposited for public access via the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo; GSM Accession nos. GSM4899 and GSM4900). Combined, a total of 27 430 tags (~29% of total tags) represented unique putative transcripts present in these two libraries. Among them 9701 (36%) transcripts were expressed in both libraries, whereas 9400 (34%) and 8329 (30%) putative transcripts were unique to the control and sperm insemination libraries, respectively. Statistical comparison of SAGE tags from each library revealed that 214 potential transcripts (0.78% [214/27 430] of total unique putative genes) were expressed at significantly different levels ($P < 0.05$), with 121 putative genes (0.44% of the total) being up-regulated and 93 (0.34% of the total) down-regulated between the sperm and control insemination libraries, respectively. To rule out the possibility that sperm RNA may have contributed to the gene expression detected in the sperm AI library, we attempted to extract RNA from 10^8 turkey spermatozoa for PCR analysis of putative genes detected by SAGE. No detectable RNA was isolated from these turkey sperm (data not shown).

To assess the representation of expressed genes relative to the overall size of each library, the accrual rate of unique tags identified was plotted as a function of the total tags sequenced across a series of random subsets of SAGE tags from within each library (Fig. 1). The accrual rate of new unique tags decreased steadily and then became more constant as the total number of tags sequenced approached 50 000. Unique tag accrual rates and patterns did not differ between the control and sperm AI SAGE libraries.

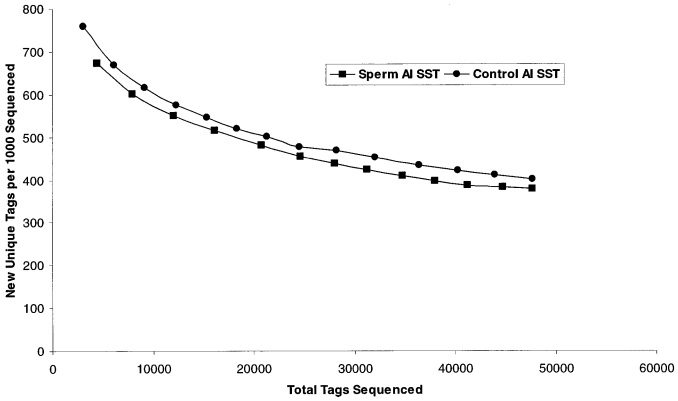

FIG. 1. SAGE tag accrual rates in sperm AI and control AI turkey SST libraries.

Identification and Annotation of Tags Representing Putative Genes

Comparing the differentially expressed ($P < 0.05$), SST-derived SAGE tag sequences against the human SAGEmap database yielded only six tentative gene annotations. In an attempt to overcome the lack of poultry-related sequence information in SAGEmap, turkey SAGE tags obtained from the sperm and control AI libraries were searched against the TIGR chicken gene index database. The initial TIGR chicken gene index database comparison yielded 490 432 records. In nearly every case, each SAGE tag corresponded with multiple tentative annotations. Filtering these data to include only those tags that were differentially ($P < 0.05$) expressed, that were represented in a library at least five times, and that were linked to a TIGR TC identifier reduced the total data set from 490 432 to 5923 records. Of the remaining 5923 records, 3648 records were tentatively annotated only as EST clones, leaving just 2275 records that annotated to a specific, definable gene. Review of these 2275 tentative annotation records for recurring themes (e.g., glucose, kinase, tubulin, etc.) enabled ''word-search''-based parsing of the data by tentative function, as represented in Table 1. Comparing the complete SST SAGE tag data set with the BBSRC Chicken EST database yielded only EST clone identifiers without more detailed gene annotations and thus provided no additional information to the TIGR database search results.

Amplification of Turkey Avidin cDNA Fragment by 5′-RACE

The SAGE tag sequence (5′-TGGCTGCTGC-3′) appeared approximately 3-fold more frequently in the sperm AI library and was tentatively annotated as *avidin* in the TIGR chicken gene index. The BLAST analysis of this tag sequence against poultry-annotated sequences in GenBank (nr and dbEST databases) significantly matched near the 3′-end of an expressed mRNA for chicken *avidin* (GenBank accession no. X05343). Sequence alignment indicated that the turkey *avidin* SAGE tag sequence mapped to 184 bases 5′ of the Poly-A tail at the 3′ most distal *Nla*III-recognition (CATG) site of the chicken *avidin* sequence. Using 5′-RACE methodology, a 412-bp cDNA fragment was amplified using the RNA representing the sperm treatment SST sample. The sequence of this turkey cDNA clone (GenBank accession no. AF545846) shared 92% nucleotide sequence homology with chicken *avidin* and thus confirmed that the SAGE tag (5′-TGGCTGCTGC-3′) represented *av-*

TABLE 1. Tentative annotations of putative genes represented by differentially expressed SAGE tags in control AI and sperm AI turkey SST SAGE libraries by comparison with the TIGR chicken gene index (version 4).

Functional grouping	SAGE tag sequence	SAGE tag count[a] Control AI[b]	Sperm AI[b]	TC identifier	Annotation
Miscellaneous	CCGGTCGCCC	13	25	TC22536	Nicotinic acetylcholine receptor, α_1 subunit (*Gallus gallus*)
	TGGCTGCTGC	138	371	TC22445	Precursor polypeptide (amino acids −24 to 128) (*G. gallus*) avidin
Cytoskeleton	GAGGAGGGGG	79	104	TC15990	Alpha tubulin (*G. gallus*)
	GTGGGGCCCA	35	59	TC20744	γ-Actin (chicken)
	TCCTCGTCAC	17	35	TC16025	α-Actin (*G. gallus*)
	TTGGAGCGAA	8	24	TC20740	β-Actin (*G. gallus*)
Metabolism	AGCCCCTATA	49	81	TC15976	Phosphoenolpyruvate carboxykinase
	TCCCCGTACA	28	54	TC23279	Similar to PIR JC5088 JC5088 pyruvate dehydrogenase (lipoamide)
	TCCCCATCTC	12	44	TC17604	Mitochondrial uncoupling protein (*G. gallus*)
	CAGCATAGAG	50	80	TC21436	Similar to GP 9652182 gb AAF91430.1 adenosine deaminase
Membrane transport	GATTTGGGGA	154	84	TC18784	Facilitative glucose transporter (*G. gallus*)
	TAGATTTATG	6	23	TC19428	Plasma membrane calcium pump (*G. gallus*)
	CAGAACCCGC	13	26	TC21280	N-type calcium channel α_{1B} subunit (*G. gallus*)
Intracellular processing	GTTGCCTGCT	8	25	TC17612	Heat shock protein 90 beta (*G. gallus*)
	ACACATTTGC	78	50	TC16564	Trans-Golgi network protease furin (*G. gallus*)

[a] SAGE tag count differences between the libraries were significant ($P < 0.05$).
[b] The control and sperm AI libraries consisted of 47 662 and 47 663 total tags, respectively.

idin mRNA expressed in turkey SST. Another SAGE tag sequence (5′-CATGGCATCCAAGG-3′) matched the 3′-end of chicken *GAPDH* in the TIGR database. This tag was not differentially expressed between the two SST SAGE libraries (61 and 78 counts in control and sperm insemination libraries, respectively). Therefore, *GAPDH* was used as a sample reference control for subsequent validation experiments based on real-time RT-PCR assays.

Confirmation of Differential Avidin Gene Expression

To confirm the differential expression of the *avidin* gene in SST mucosal epithelium revealed by SAGE analysis, real-time PCR was performed using SST RNA from the sperm and control AI treatments. In this assay, differences in gene expression are reflected by differences in C_T whereby the shorter the C_T, the higher the concentration of the initial specific mRNA template in the sample. The C_T for *avidin* amplification was significantly shorter ($P = 9.62E-11$) in the sperm-treated versus control SST samples. No significant difference was found in the C_T values for *GAPDH* amplification (Table 2). A standard-curve analysis was then performed to rule out the possibility that differential expression of *avidin* between the treatments was an artifact of the real-time PCR reaction conditions. This analysis revealed AE values of 1.46 and 1.70 for *GAPDH* and *avidin*, respectively. Using these values and the C_T values obtained for both genes, the difference in *avidin* expression between the sperm and control insemination samples was determined to be approximately 2.8-fold (Table 3).

TABLE 2. Real-time RT-PCR analysis of *avidin* expression between control AI and sperm AI SST.[a]

	C_T (mean ± SEM) Sham-inseminated	AI	*P*
GAPDH	34.63 ± 0.76	34.61 ± 0.33	0.9466
Avidin	25.88 ± 0.23	23.45 ± 0.27	9.62E-11

[a] Results represent three independent experiments. Each PCR reaction was performed in quadruplicate within each experiment.

DISCUSSION

We constructed two turkey SAGE libraries comprised of 95 325 total tags to represent and compare differential gene expression within sperm and control inseminated SST. These libraries represented 27 430 (~29% of total tags) unique putative transcripts. This level of representation of putative unique genes is consistent with SAGE analyses reported for other model systems [21, 29, 30]. It is interesting to note that only 214 of the 27 430 (~1%) putative genes represented between the sperm inseminated and control SAGE libraries were differentially expressed between these treatments; the relative frequencies of the vast majority of tags were beneath the $P < 0.05$ statistical threshold. This 1% of differential gene expression in turkey SST is similar to SAGE results comparing developing mouse forelimbs and hindlimbs, in which 1% (317/36 300) of the unique transcripts were differentially expressed [30]. The lack of detectable RNA in turkey sperm rules out the possibility that sperm RNA may have contributed to the SAGE results obtained from the sperm inseminated SST. These data confirm our initial hypothesis that differential gene expression occurs in the SST mucosal epithelium by 48 h after insemination with sperm and thereby raises the compelling possibility that the spermatozoa themselves induce specific gene expression events required for prolonged sperm storage in the SST.

A complete SAGE analysis of the human transcriptome

TABLE 3. Standard-curve real-time PCR analysis comparing *avidin* expression between control AI and sperm AI SST.

Experiment	C_T (*GAPDH*) Sham-inseminated	AI	C_T (*avidin*) Sham-inseminated	AI	Fold induction
1	40.44	42.96	30.53	29.98	3.47
2	40.76	42.44	31.22	30.34	3.04
3	43.37	40.89	32.17	29.20	1.92
					2.81 (0.80)[a]

[a] Mean (SD).

has been estimated to require sequencing of approximately 650 000 tags [31]. However, in practice, accurate and efficient representation of most transcriptomes (≥100 mRNA copies) is obtained after analysis of far fewer tags, most commonly in the range of 50 000 SAGE tags [21, 31, 32]. In silico analyses indicate that 50 000 SAGE tags sufficiently account for more than 98% of significant ($P < 0.05$) 2-fold differences at the 0.1% level of expression (i.e., 1 tag per 1000 tags sequenced) [33]. Therefore, we established an initial target size of approximately 50 000 tags for each of the SST SAGE libraries. As depicted in Figure 1, the new tag accrual rates tapered off and reached a ''steady state'' after 35 000 tags were sequenced in the sperm and control AI SST libraries. This steady state suggests that new tags were being sequenced at the same rate as previously unique tags became redundant. More than 76% of the total tags were represented at least twice in each final SST library (data not shown). Thus, the approximately 50 000 tags in each of the current turkey SST SAGE libraries probably represent the most common, along with many of the rare, unique genes expressed in the SST transcriptome.

It was not surprising that the initial BLAST comparison of the turkey SST SAGE tag sequences against the human SAGEmap database yielded so few tentative annotations. This low yield of tentative annotations probably reflects existing sequence variation between avian and human (mammalian) species at the 3′-ends of mRNA transcripts from which the SAGE tags originated. Comparing the turkey SAGE tag sequences against the TIGR chicken gene index database yielded more results; however, only approximately 0.5% (2275/490 432) of these results were associated with tentative functional annotations. One should view these initial database comparisons as a baseline for tentatively annotating the genes represented by the SAGE tags sequenced in the present study. As more chicken genomic and cDNA sequence data become available in the future, serial updating of the database search queries established in the present study will almost certainly yield more extensive and detailed tag-to-gene annotations. It is worth noting again the results of these BLAST analyses were obtained by comparing the turkey SAGE sequences to expressed gene sequences in the TIGR chicken gene index due to the lack of turkey sequence information. Because of the potential sequence variation between chickens and turkeys and the lack of an automated analysis tool to determine SAGE tag position relative to the matched chicken sequence (i.e., 3′ vs. 5′ proximity), these tentative BLAST matches require further validation to confirm SAGE tag annotation and gene identity.

Bioinformatic analyses of the SST SAGE tag libraries against both the TIGR chicken gene index and GenBank nr databases suggested that one of these differentially expressed SAGE tags represented the *avidin* gene. Extending and amplifying this tag with 5′-RACE combined with nucleotide sequence analysis of the resulting cDNA clone confirmed that this tag indeed represented the turkey *avidin* gene. To our knowledge, turkey *avidin* had not been cloned previously. Quantitative real-time PCR analysis corroborated the SAGE analysis data, indicating that *avidin* mRNA expression is up-regulated approximately 3-fold in SST following sperm insemination.

The role, if any, that *avidin* may play in mediating prolonged sperm storage within the SST is not known. Avidin is a major turkey egg-white protein that binds and sequesters biotin within the egg [34]. It is possible that increased *avidin* expression in sperm-containing SST may provide the sperm that are resident within those SST with a nutritional source of biotin or related vitamins. Induction of *avidin* expression is commonly employed as a marker of progesterone activity in chick oviduct [35, 36]. Thus, increased *avidin* expression in sperm SST may also reflect a potential linkage between prolonged sperm storage and release and progesterone fluctuations in laying hens. Our laboratory is currently performing follow-up studies to localize *avidin* expression in the SST epithelium and verify its potential up-regulation in response to sperm at the cellular level.

The expression levels of multiple cytoskeletal protein genes appeared to be increased in the sperm AI SST (Table 1). The SST epithelial cells possess an intricate F-actin-rich terminal web that may play a role in mediating contractile activity within the SST epithelium [37]. It is possible that the increased expression of cytoskeletal protein genes in the sperm AI SST may be indicative of increased contractility to facilitate sperm selection, storage, and then release from the SST.

Several strategies have been developed to elucidate modulated gene expression. Most of these strategies require the existence of extensively characterized cDNA libraries to generate EST information [38, 39]. This is especially critical for microarray hybridization and PCR-based subtractive cloning. Even though this information greatly facilitates the annotation of SAGE tags, we demonstrated that species-specific EST sequences are not required to perform the statistical analysis of SAGE data that leads to the identification of some genes important to a physiological event of interest. The analytical steps by which the differential expression of the *avidin* gene was confirmed represent a useful paradigm for taking SAGE tag data from a relatively uncharacterized model system (i.e., turkey SST) to identify differentially expressed genes, to confirm identity by nucleotide sequence homology and 5′-RACE, and to validate differential expression through independent qualitative and quantitative analyses (e.g., real-time RT-PCR). Our laboratory is currently validating the tentative annotation and expression levels of an extended group of differentially expressed SST SAGE tags to establish a working list of candidate genes to be progressed together in specific follow-up experiments investigating avian sperm storage mechanisms. We propose that the present SAGE analysis paradigm could be extremely useful to investigate differential gene expression on a transcriptome-wide level in a wide range of otherwise poorly characterized model systems.

The overall aim of the present study was to establish and characterize the range of genes expressed in turkey SST following insemination with or without sperm to identify specific genes that may enable and regulate prolonged sperm storage. Our SAGE-based strategy achieved this aim by yielding quantitative assessments of gene expression independent of original mRNA sequence knowledge and by providing the capacity for high-throughput and cost-efficient analysis on a transcriptome-wide level. In addition, these SAGE libraries provided absolute transcript numbers in a digital format that can be adapted to provide statistical comparisons of data from multiple laboratories [20]. It must be noted that mechanisms other than differential gene expression (e.g., posttranslational modifications, relocation of proteins in SST epithelial cells, etc.) also likely are important factors in SST-mediated sperm storage. Proteomic analyses of SST-mediated sperm storage will be critical to elucidate these additional mechanisms. The SAGE analyses conducted in the present study are only the beginning of an iterative process of hypothesis development, testing, and

refinement that will ultimately yield new insights regarding the physiologic mechanisms that underlie prolonged sperm storage in the avian oviduct.

ACKNOWLEDGMENTS

The authors thank Jane Garlow for turkey sperm RNA analyses, Tina Sphon for assistance in DNA sequencing, Larry Shade and Jill Philpot for assistance in bioinformatics, and Drs. Laura King and Murray Bakst for assistance in turkey SST anatomy and collection.

REFERENCES

1. Bakst MR. Oviductal sperm selection, transport, and storage in poultry. Poult Sci Rev 1994; 5:117–143.
2. Christensen VL. Effect of insemination intervals on oviductal sperm storage in turkeys. Poult Sci 1981; 60:2150–2156.
3. Hardy MP, Dent JN. Transport of sperm within the cloaca of the female red-spotted newt. J Morphol 1986; 190:259–270.
4. Hoffman LH, Wimsatt WA. Histochemical and electron microscopic observations on the sperm receptacles in the garter snake oviduct. Am J Anat 1972; 134:71–96.
5. Sever DM, Ryan TJ. Ultrastructure of the reproductive system of the black swamp snake (*Seminatrix pygaea*): part I. Evidence for oviductal sperm storage. J Morphol 1999; 241:1–18.
6. Sever DM. Female sperm storage in amphibians. J Exp Zool 2002; 292:165–179.
7. Bou-Resli MN, Bishay LF, Al-Zaid NS. Observations on the fine structure of the sperm storage crypts in the lizard *Acanthodactylus scutellatus hardyi*. Arch Biol 1981; 92:287–298.
8. Gist DH, Jones JM. Sperm storage within the oviduct of turtles. J Morphol 1989; 199:379–384.
9. Brillard JP. Sperm storage and transport following natural mating and artificial insemination. Poult Sci 1993; 72:923–928.
10. Bakst MR. Structure of the avian oviduct with emphasis on sperm storage in poultry. J Exp Zool 1998; 282:618–626.
11. King LM, Brillard JP, Garrett WM, Bakst MR, Donoghue AM. Segregation of spermatozoa within sperm storage tubules of fowl and turkey hens. Reproduction 2002; 123:79–86.
12. Christensen L. Diluents, dilution and storage of turkey semen for six hours. In: Bakst MR, Wishart GH (eds.), First International Symposium on the Artificial Insemination of Poultry. Savoy, IL: Poultry Science Association; 1995:90–106.
13. Wishart GJ. Physiological changes in fowl and turkey spermatozoa during in vitro storage. Br Poult Sci 1989; 22:443–454.
14. Bakst MR. Oviductal sperm storage in poultry: a review. Reprod Fertil Dev 1993; 5:595–599.
15. Holm L, Wishart GJ. The effect of pH on the motility of spermatozoa from chicken, turkey and quail. Anim Reprod Sci 1998; 54:45–54.
16. Wishart GJ, Wilson YI. Temperature-dependent inhibition of motility in spermatozoa from different avian species. Anim Reprod Sci 1999; 57:229–235.
17. Holm L, Dkwall H, Wishart GJ, Ridderstråle Y. Localisation of calcium and zinc in the sperm storage tubules from chicken, quail and turkey using x-ray microanalysis. J Reprod Fertil 2000; 118:331–336.
18. Holm L, Ridderstråle Y. Localization of carbonic anhydrase in the sperm-storing regions of the turkey and quail oviduct. Histochem J 1998; 30:481–488.
19. Yoshimura Y, Koike K, Okamoto T. Immunolocalization of progesterone and estrogen receptors in the sperm storage tubules of laying and diethylstilbestrol-injected immature hens. Poult Sci 2000; 79:94–98.
20. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science 1995; 270:484–487.
21. Velculescu VE, Vogelstein B, Kinzler KW. Analysing uncharted transcriptomes with SAGE. Trends Genet 2000; 16:423–425.
22. Bakst MR. Fate of fluorescent stained sperm following insemination: new light on oviductal sperm transport and storage in the turkey. Biol Reprod 1994; 50:987–992.
23. King LM, Brillard JP, Bakst MR, Donoghue AM. Isolation of sperm storage tubules from the uterovaginal junction mucosa of the turkey. Poult Sci 1999; 78:1044–1047.
24. Audic S, Claverie JM. The significance of digital gene expression profiles. Genome Res 1997; 7:986–995.
25. Lash AE, Tolstoshev CM, Wagner L, Schuler GD, Strausberg RL, Riggins GJ, Altschul GD. SAGEmap: a public gene expression resource. Genome Res 2000; 10:1051–1060.
26. Quackenbush J, Cho J, Lee D, Liang F, Holt I, Karamycheva S, Parvizi B, Pertea G, Sultana R, White J. The TIGR gene indices: analysis of gene transcript sequences in highly sampled eukaryotic species. Nucleic Acid Res 2001; 29:159–164.
27. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215:403–410.
28. Vanden Berg A, van der Leij J, Poppema S. Serial analysis of gene expression: rapid RT-PCR analysis of unknown SAGE tags. Nucleic Acids Res 1999; 27(e17):i–iii.
29. Menssen A, Hermeking H. Characterization of the c-MYC-regulated transcriptome by SAGE: identification and analysis of c-MYC target genes. Proc Natl Acad Sci U S A 2002; 99:6274–6279.
30. Margulies EH, Kardia SLR, Innis JW. A comparative molecular analysis of developing mouse forelimbs and hindlimbs using serial analysis of gene expression (SAGE). Genome Res 2001; 11:1686–1698.
31. Velculescu VE, Madden SL, Zhang L, Lash AE, Yu J, Rago C, Lal A, Wang CJ, Beaudry GA, Ciriello KM, Cook BP, Dufault MR, Ferguson AT, Gao Y, He TC, Hermeking H, Hiraldo SK, Hwang PM, Lopez MA, Luderer HF, Mathews B, Petroziello JM, Polyak K, Zawel L, Kinzler KW. Analysis of human transcriptomes. Nat Genet 1999; 23:387–388.
32. Anisimov SV, Tarasov KV, Tweedie D, Stern MD, Wobus AM, Bohler KR. SAGE identification of gene transcripts with profiles unique to pluripotent mouse R1 embryonic stem cells. Genomics 2002; 79:169–176.
33. Man MZ, Wang X, Wang Y. Power_SAGE: comparing statistical tests of SAGE experiments. Bioinformatics 2000; 16:953–959.
34. White HB III. Biotin-binding proteins and biotin transport to oocytes. Ann N Y Acad Sci 1985; 447:202–211.
35. Niemelä AO, Elo HA. Effects of oestradiol-17β and diethylstilboestrol on progesterone-induced protein (avidin) production in chick oviduct: evidence for differences in the actions of steroidal and non-steroidal oestrogens. J Endocrinol 1983; 96:465–469.
36. Hora J, Gosse B, Rasmussen K, Spelsberg TC. Estrogen regulation of the biological activity of the avian oviduct progesterone receptor and its ability to induce avidin. Endocrinology 1986; 119:1118–1125.
37. Freedman S, Akuffo V, Bakst MR. New evidence for the innervation of the sperm storage tubules in the turkey (*Meleagris gallopavo*). Reproduction 2001; 121:809–814.
38. Marra MA, Hillier L, Waterston RH. Expressed sequence tags—Establishing bridges between genomes. Trends Genet 1998; 14:4–7.
39. Kozian DH, Kirschbaum BJ. Comparative gene-expression analysis. TIBTECH 1999; 17:73–78.